

- <http://zakon0.rada.gov.ua/laws/show/326-93-p> (accessed: 03.03.2020).
8. *Pro standartyzatsiyu* [About standardization] : zakon Ukrainy vid 05.06.2014 no. 1315–VII URL: <http://zakon3.rada.gov.ua/laws/show/1163-2014-p> (accessed: 03.03.2020).
 9. *Pro vyznachennya derzhavnoho pidpryyemstva, yake vykonuye funktsiyi natsional'noho orhanu standartyzatsiyi* [On the definition of a state-owned enterprise that performs the functions of a national standardization body] : rozporядzhennya Kabinetu Ministriv Ukrainy vid 26.11.2014 no. 1163-r URL: <http://zakon3.rada.gov.ua/laws/show/1163-2014-p> (accessed: 03.03.2020).
 10. *Poryadok vedennya vidomchykh klasyfikatoriv informatsiyi z pytan' derzhavnoyi mytnoyi spravy, yaki vykorystovuyut'sya u protsesi oformlennya mytnykh deklaratsiy* [Procedure for keeping departmental classifiers of information on state customs matters used in the process of customs declaration processing] URL: <http://zakon0.rada.gov.ua/laws/show/z0989-12> (accessed: 03.03.2020).
 11. K-84.2-37552598-003:2018 *Klasyfikator dokumentiv strakhovoho fondu dokumentatsiyi Ukrainy* [Classifier of documents of insurance fund of documentation of Ukraine]. Kharkiv: NDI mikrohrafiyi, 2018. 77 p.
 12. DK 009:2010 *Derzhavnyy klasyfikator Ukrainy. Klasyfikatsiya vydiv ekonomichnoyi diyal'nosti* [State Classifier of Ukraine. Classification of economic activities]. Kyiv: Derzhspozhyvstandart Ukrainy, 2010. 42 p.
 13. *Statistical Classification of Economic Activities in the European Community*. URL: <http://ec.europa.eu/eurostat/documents/3859598/5902521/KS-RA-07-015-EN.PDF> (accessed: 03.03.2020).

Надійшло (received) 21.05.2020

Відомості про авторів / Сведения об авторах / About the Authors

Ситник Наталія Леонідівна – Науково-дослідний, проектно-конструкторський та технологічний інститут мікрографії (НДІ мікрографії), молодший науковий співробітник; м. Харків, Україна; e-mail: nataly.sytnik1977@gmail.com

Виноградова Ольга Євгенівна – Науково-дослідний, проектно-конструкторський та технологічний інститут мікрографії (НДІ мікрографії), провідний інженер із стандартизації; м. Харків, Україна; e-mail: olga.kroots@gmail.com

Тягун Тетяна Володимирівна – Науково-дослідний, проектно-конструкторський та технологічний інститут мікрографії (НДІ мікрографії), провідний інженер-проектувальник; м. Харків, Україна; e-mail: rusikesh8@gmail.com

Мазничко Андрій Борисович – Науково-дослідний, проектно-конструкторський та технологічний інститут мікрографії (НДІ мікрографії), провідний інженер-програміст; м. Харків, Україна; e-mail: a.maznychko.ndi_m@arch.gov.ua

Сытник Наталья Леонидовна – Научно-исследовательский, проектно-конструкторский и технологический институт микрографии (НИИ микрографии), младший научный сотрудник; г. Харьков, Украина; e-mail: nataly.sytnik1977@gmail.com

Виноградова Ольга Евгеньевна – Научно-исследовательский, проектно-конструкторский и технологический институт микрографии (НИИ микрографии), ведущий инженер по стандартизации; г. Харьков, Украина; e-mail: olga.kroots@gmail.com

Тягун Татьяна Владимировна – Научно-исследовательский, проектно-конструкторский и технологический институт микрографии (НИИ микрографии), ведущий инженер-проектировщик; г. Харьков, Украина; e-mail: rusikesh8@gmail.com

Мазничко Андрей Борисович – Научно-исследовательский, проектно-конструкторский и технологический институт микрографии (НИИ микрографии), ведущий инженер-программист; г. Харьков, Украина; e-mail: a.maznychko.ndi_m@arch.gov.ua

Sytnyk Nataliia Leonidivna – project-design and technological institute of micrography (SRI micrography), Research Assistant; Kharkiv, Ukraine; e-mail: nataly.sytnik1977@gmail.com

Vynogradova Olga Yevheniivna – Scientific-research, project-design and technological institute of micrography (SRI micrography), leading standardization engineer; Kharkiv, Ukraine; e-mail: olga.kroots@gmail.com

Tiahun Tetiana Volodymyrivna – Scientific-research, project-design and technological institute of micrography (SRI micrography), leading design engineer; Kharkiv, Ukraine; e-mail: rusikesh8@gmail.com

Maznychko Andrii Borysovych – Scientific-research, project-design and technological institute of micrography (SRI micrography), leading software engineer; Kharkiv, Ukraine; e-mail: a.maznychko.ndi_m@arch.gov.ua

УДК 681.3.06

DOI: 10.20998/2079-0023.2020.01.07

А. Ю. БАРАНЦЕВ, Н. М. КЛИМЕНКО, І. А. ШЕВЧЕНКО

ПРОБЛЕМНІ ПИТАННЯ ЗБЕРЕЖЕННЯ ТА ОБРОБКИ ВЕЛИКИХ ДАНИХ ДЛЯ ЗАБЕЗПЕЧЕННЯ ОБЛІКУ ТА ПІДГОТОВКИ ДО ЕКСПОНУВАННЯ ЕЛЕКТРОННИХ ФОНДІВ АРХІВНОЇ УСТАНОВИ В ПУБЛІЧНИХ ЕЛЕКТРОННИХ МЕРЕЖАХ

Великі темпи зростання обсягу даних в інформаційних технологіях є невід'ємною складовою сучасного суспільства. Використання інформаційних технологій в архівній справі дозволяє підвищити ефективність надання державними архівними установами послуг, забезпечити популяризацію їхньої діяльності на новому якісному рівні шляхом застосування сучасних форм публікації та експонування архівних документів в електронному вигляді. Так, завдяки використанню інформаційних технологій користувачі Федерального архіву Німеччини мають можливість здійснити повнотекстовий пошук у колекції документів архіву, переглянути в зручній формі результати пошуку (зокрема у вигляді цифрованих зображень документів), а також, за потреби, замовити та сплатити певні додаткові послуги (наприклад, копію документа тощо). З аналогічною метою за підтримки ЮНЕСКО та Бібліотеки Конгресу США було створено вебсайт Світової цифрової бібліотеки. Метою дослідження є визначення належної програмної інфраструктури та об'єктної моделі організації даних, необхідних для

© А. Ю. Баранцев, Н. М. Клименко, І. А. Шевченко, 2020

забезпечення державного обліку документів Національного архівного фонду (облікової та вторинної архівної інформації) з метою створення якісної електронної експозиції та забезпечення користувачів можливістю вільного доступу до неї.

Ключові слова: цифровізація, Національний архівний фонд, архів, бібліотека, об'єктна модель, експозиція, клієнт, програмне забезпечення, сервер, Центральний державний науково-технічний архів України.

А. Ю. БАРАНЦЕВ, Н. Н. КЛИМЕНКО, И. А. ШЕВЧЕНКО

ПРОБЛЕМНЫЕ ВОПРОСЫ СОХРАНЕНИЯ И ОБРАБОТКИ БОЛЬШИХ ДАННЫХ ДЛЯ ОБЕСПЕЧЕНИЯ УЧЕТА И ПОДГОТОВКИ К ЭКСПОНИРОВАНИЮ ЭЛЕКТРОННЫХ ФОНДОВ АРХИВНОГО УЧРЕЖДЕНИЯ В ПУБЛИЧНЫХ ЭЛЕКТРОННЫХ СЕТЯХ

Большие темпы роста объема данных в информационных технологиях являются неотъемлемой составляющей современного общества. Использование информационных технологий в архивном деле позволяет повысить эффективность предоставления государственным архивными учреждениями услуг, обеспечить популяризацию их деятельности на новом качественном уровне путем применения современных форм публикации и экспонирования архивных документов в электронном виде. Так, благодаря использованию информационных технологий пользователи Федерального архива Германии имеют возможность осуществить полнотекстовый поиск в коллекции документов архива, просмотреть в удобной форме результаты поиска (в частности в виде цифровых изображений документов), а также, при необходимости, заказать и оплатить определенные дополнительные услуги (например, копию документа и т.д.). С аналогичной целью при поддержке ЮНЕСКО и Библиотеки Конгресса США был создан веб-сайт Всемирной цифровой библиотеки. Целью исследования является определение надлежащей программной инфраструктуры и объектной модели организации данных, необходимых для обеспечения государственного учета документов Национального архивного фонда (учетной и вторичной архивной информации) с целью создания качественной электронной экспозиции и обеспечения пользователей возможностью свободного доступа к ней.

Ключевые слова: цифровизация, Национальный архивный фонд, архив, библиотека, объектная модель, экспозиция, клиент, программное обеспечение, сервер, Центральный государственный научно-технический архив Украины.

A. Y. BARANTSEV, N. M. KLYMENKO, I. A. SHEVCHENKO

PROBLEMS ISSUES ON SAVING AND PROCESSING OF BIG DATA TO ENSURE THE ACCOUNTING AND PREPARATION FOR EXPOSURE OF ELECTRONIC FUNDS OF ARCHIVE INSTITUTION IN PUBLIC ELECTRONIC NETWORKS

Big data growth in information technology is an integral part of today's society. The use of information technologies in archival affairs allows to increase the efficiency of providing state archival institutions services, to ensure promotion of their activity at a new quality level by applying modern forms of publication and exhibiting archival documents in electronic form. So, thanks to using of information technology, users of the Federal Archives of Germany have the opportunity to search in full text search in the archive documents collection, to view in a comfortable form search results (in particular in the form of digital images of documents), and, if necessary, to order and pay for certain additional services (e.g. a copy of the document, etc.). For the same purpose, the World Digital Library website was created with the support of UNESCO and the Library of Congress. The purpose of the study is to determine the appropriate software infrastructure and an object model for organizing data needed to ensure state registration of documents of the National Archival Fund (accounting and secondary archival information) in order to create high-quality electronic exposure and provide users with free access to it.

Keywords: digitalization, national archival fund, archive, library, object model, exposition, client, software, server, Central State Scientific and Technical Archive of Ukraine.

Вступ. Великі темпи зростання обсягу даних в інформаційних технологіях (далі – ІТ) є невід'ємною складовою сучасного суспільства. Використання ІТ в архівній справі дозволяє підвищити ефективність надання державними архівними установами послуг, забезпечити популяризацію їхньої діяльності на новому якісному рівні шляхом застосування сучасних форм публікації та експонування архівних документів в електронному вигляді. Так, завдяки використанню ІТ користувачі Федерального архіву Німеччини мають можливість здійснити повнотекстовий пошук у колекції документів архіву, переглянути в зручній формі результати пошуку (зокрема у вигляді цифрованих зображень документів), а також, за потреби, замовити та сплатити певні додаткові послуги (наприклад, копію документа тощо) [1].

З аналогічною метою за підтримки ЮНЕСКО та Бібліотеки Конгресу США було створено вебсайт Світової цифрової бібліотеки [2].

Мета дослідження. Забезпечити архівні установи України можливістю ефективно вести облік та накопичення вторинної архівної інформації для надання вільного доступу до неї користувачам.

Постановка проблеми. На цей час з метою забезпечення прав громадян на вільний доступ до відомостей, потрібних для науково-дослідної, інформаційно-довідкової та іншої діяльності, державними архівними установами розроблено довідкові апарати у формі електронних баз даних (далі – БД) та створено офіційні вебсайти. На вебсайтах експонуються документи Національного архівного фонду (далі – НАФ) у вигляді тематичних експозицій. Це, наприклад, перелік електронних експозицій на вебсайті Державної архівної служби України [3] або електронна експозиція на вебсайті Центрального державного науково-технічного архіву України (далі – ЦДНТА) [4]. Такі електронні експозиції містять документи, підібрані фахівцями архівної установи за певними критеріями. Основна відмінність цих експозицій, наприклад, від експозиції Федерального архіву Німеччини полягає в тому, що в ній представлені всі документи архіву, вторинна архівна інформація та надано потужний механізм пошуку. При такому підході до побудови електронної експозиції критерії відбору документів визначає користувач, що робить її придатною для науково-дослідної та інформаційно-довідкової діяльності користувачів.

Таким чином, однією з вимог до сучасної електронної експозиції є експонування архівних документів в електронному вигляді разом з обліковою та вторинною архівною інформацією.

Таким чином, однією з вимог до сучасної електронної експозиції є експонування архівних документів в електронному вигляді разом з обліковою та вторинною архівною інформацією.

Одним з пріоритетів Державної архівної служби України на 2020 рік є цифровізація всіх процесів архівної справи, діловодства та державної системи страхового фонду документації. Це у свою чергу несе за собою необхідність зберігати та обробляти великі масиви даних для інформаційного забезпечення архівної справи.

Виклад основного матеріалу. Багато сучасних організацій, що використовують великі дані, не створюють належної інфраструктури, яка змогла б забезпечити надійне зберігання величезних масивів даних. Це, на наш погляд, є неправильним та містить у собі низку проблем:

- це просто багато даних (> 10 Тбайт);
- це настільки великі таблиці, що їх неможливо відкрити та проаналізувати в Excel;
- це неструктуровані дані із різних джерел, різного об'єму.

Але ми зупинимось на класичному визначенні.

Великі дані (англ. Big Data) в інформаційних технологіях – набори інформації (як структурованої, так і неструктурованої) настільки великих розмірів, що традиційні способи та підходи (здебільшого засновані на рішеннях класу бізнесової аналітики та системах керування базами даних (далі – СКБД)) не можуть бути застосовані до них. Альтернативне визначення називає великими даними феноменальне прискорення нагромадження даних та їх ускладнення.

Окрім фізичного об'єму даних, існує так звана

3-V теорія, яка підкреслює складність задачі обробки і аналізу цих даних (набір ознак VVV – volume, velocity, variety, – фізичний об'єм, швидкість приросту даних і необхідність їх швидкої обробки, здатність обробляти дані різних типів). Теорію було висунуто компанією Meta Group у 2001 році з метою наголосити на рівній значимості управління даними по всіх трьох аспектах.

Розглянемо кілька прикладів, де використовуються великі дані:

- логи сайтів, які накопичують і зберігають службову інформацію про відвідувачів;
- інформація про операції клієнтів банку;
- оцифровані книги та документи;
- інформація про покупки у великих торговельних мережах;
- сигнали GPS від автомобілів для транспортувальної компанії;
- інформація про абонентів мобільних операторів;
- та багато інших [5].

Державними архівними установами ведуться роботи з оцифрування архівних документів, а державний облік документів НАФ здійснюється також і в електронному вигляді. Але цього недостатньо для створення якісної електронної експозиції з кількох причин: по-перше, архівні електронні документи зберігаються переважно у форматах TIFF або PDF [6]. Ці формати не популярні у WEB. Так, растрові зображення у форматі TIFF не відображаються такими розповсюдженими Інтернет-оглядачами, як Mozilla Firefox, Google Chrome та Opera без встановлення додаткових компонентів. Вміст файлів у форматі PDF

взагалі неможливо інтегрувати до вмісту вебсторінки. Тому завдання зі створення сучасної електронної експозиції вимагає інших форматів растрових зображень. По-друге, об'єм вмісту вторинної архівної інформації, що підлягає експонуванню, у загальному випадку може не збігатися з її повним об'ємом. Звісно, рішення про це приймається фахівцями архівної установи, які готують експозицію.

Отже, для успішного вирішення завдань зі створення інформаційних електронних ресурсів і організації доступу до них (у комп'ютеризованих читальних залах, у режимі віддаленого доступу через мережу Інтернет тощо) необхідно вести в електронному вигляді окремий облік вмісту, який може бути експоновано. Такий облік необхідно вести на захищених апаратних засобах, принаймні без доступу до публічних мереж. Це надає можливість попередньо підготувати експозицію або її частину, а в публічні мережі перенесити тільки перевірені, логічно об'єднані дані.

Завданнями обліку є:

- накопичення вторинної архівної інформації;
- накопичення оцифрованих документів та розпізнаних текстів;
- перетворення графічних та текстових форматів у форму, придатну для електронної експозиції;
- систематизація накопиченої інформації за допомогою сучасних інформаційних технологій;
- відбір документів та обсягу вторинної архівної інформації, які мають бути опубліковані;
- експорт вмісту для експонування у вигляді, придатному для автоматизованого розміщення на комп'ютерних засобах у публічних електронних мережах.

Для виконання цих завдань було запропоновано спеціалізоване програмне забезпечення «Електронні фонди архіву» (далі – ПЗ Елфа), яке для ефективного ведення обліку та систематизації накопиченої інформації має низку особливостей, а саме:

- оригінальну структуру БД, яка дозволяє представити структуру облікових одиниць у вигляді графа;
- повнотекстовий пошук українською та російською мовами, що дозволяє вести пошук не тільки за іменами документів, а й за їх вмістом;
- вбудовану технологію електронного маркера E-Colorer [7], яка дозволяє наочно виділити результати пошуку на растрових зображеннях цифрових копій;
- модульний принцип побудови, що дозволяє враховувати особливості предметної області конкретної архівної установи без втручання в модулі ядра.

Для обліку застосовується об'єктно-реляційна СКБД PostgreSQL, яка надає безліч можливостей, досить надійна і має хороші характеристики за продуктивністю. Вона працює практично на всіх UNIX-платформах, включаючи UNIX-подібні системи, такі як FreeBSD і Linux. Її можна застосовувати на Windows NT Server і Windows 2000 Server, а для розробки годяться навіть такі системи Microsoft для робочих станцій, як ME. Крім того, PostgreSQL вільно поширюється і має відкритий вихідний код.

PostgreSQL вигідно відрізняється від багатьох інших СКБД. Вона володіє практично всіма можливостями, які є в інших БД (комерційних або Open Source), а також деякими додатковими.

Кожна версія перевіряється дуже ретельно, бета-версії проходять як мінімум місячне тестування. Завдяки численній спільноті користувачів і відкритому доступу до вихідного коду помилки виправляються дуже швидко.

Продуктивність цієї СКБД також зростає від версії до версії, і останні атестації показують, що при певних умовах вона не поступається комерційним продуктам. Деякі системи, що володіють не таким повним набором можливостей, перевершують PostgreSQL у продуктивності, але за рахунок втрати функціональності. Для досить простих додатків таку роль відіграє БД, що складається з плоских файлів.

Однією із сильних сторін PostgreSQL є її архітектура. Як і багато комерційних СКБД, PostgreSQL може застосовуватися в середовищі клієнт-сервер, що надає масу переваг як користувачам, так і розробникам.

Основа PostgreSQL становить серверний процес БД. Він виконується на одному сервері.

Доступ з додатків до даних бази здійснюється за допомогою процесу БД. Клієнтські програми не можуть отримати доступ до даних самостійно, навіть якщо вони працюють на тому ж комп'ютері, на якому виконується серверний процес.

Такий поділ клієнтів і сервера дозволяє побудувати розподілену систему. Можна відокремити клієнтів від сервера за допомогою мережі й розробляти клієнтські програми в середовищі, зручному для користувача. Наприклад, можна реалізувати БД під UNIX і створити клієнтські програми, які будуть працювати в системі Microsoft Windows.

Кілька клієнтів приєднуються до сервера по мережі. PostgreSQL орієнтований на протокол TCP / IP – це може бути локальна мережа або Інтернет. Кожен клієнт з'єднується з основним серверним процесом БД, який створює новий серверний процес спеціально для обслуговування запитів на доступ до даних конкретного клієнта.

Завдяки тому, що маніпулювання даними зосереджено на сервері, СКБД не доводиться контролювати численних клієнтів, які отримують доступ у спільно використовуваний каталог сервера, і PostgreSQL може підтримувати цілісність даних навіть при одночасному доступі великої кількості користувачів.

Архітектура «клієнт-сервер» робить можливим розподіл праці. Машина-сервер добре підходить для зберігання і управління доступом до великих обсягів даних, вона може використовуватися як надійний репозитарій. Для клієнтів можуть бути розроблені складні графічні додатки. Як альтернативу можна розглядати зовнішній інтерфейс на основі мережі Інтернет, який надавав би доступ до даних і повертав результат у вигляді вебсторінок у стандартний веббраузер, при цьому не потрібно було б жодного додаткового клієнтського програмного забезпечення [8].

Для забезпечення доступу користувачів до БД віддаленого доступу та документів НАФ певної архівної установи і довідкового апарату до них потрібними є:

- файловий сервер для тимчасового зберігання замовлених цифрових копій документів НАФ;
- сервер БД віддаленого доступу;
- сервер вебзастосування, за допомогою якого користувачі здійснюватимуть доступ до документів НАФ та довідкового апарату до них.

Для більшого розуміння розглянемо триланкову або по-іншому – трирівневу архітектуру систем БД, пов'язану із забезпеченням спільної роботи користувачів. Дана архітектура припускає наявність додаткового сервера додатків, який проводить попередню обробку запитів клієнтів, формує запити до сервера БД і обробляє отримані результати перед відправкою їх клієнту.

У триланковій архітектурі велика частина логіки додатка перенесена з клієнта на сервер, і завдання клієнтського додатка зводяться, в основному, до реалізації призначеного для користувача інтерфейсу і подання результатів. З'являється можливість централізовано вносити зміни в алгоритми бізнес-логіки, реалізовані в програмному забезпеченні сервера додатків. Також у цій архітектурі тільки сервер додатків може підключатися до сервера БД. Це усуває проблему підтримки невикористовуваних з'єднань і є більш переважним з точки зору безпеки. Недоліком багаторівневої архітектури «клієнт-сервер» є складність розробки подібних рішень [9].

Для нашої мети сервер вебзастосувань розташовується на виділеному сервері певної архівної установи, на ньому знаходиться й сайт.

Виділений сервер – це, зазвичай, орендований у хостинг-провайдера звичайний серверний комп'ютер, тому його потужність у порівнянні з віртуальним сервером може бути вищою в рази. Доступ до управління відбувається за допомогою віддаленого з'єднання. Операційну систему у виділеного сервера встановлюють самостійно. Такий сервер має широкий діапазон налаштувань, потрібне для роботи додаткове програмне забезпечення може бути встановлене самостійно. Усе це дозволяє налаштувати сервер для конкретних завдань.

Таке розташування вебсервера у хостинг-провайдера надає нам кілька суттєвих переваг, а саме:

- відсутність витрат значних фінансових ресурсів на придбання власного серверного обладнання;
- відсутність витрат на оренду або придбання спеціального приміщення для розміщення серверів;
- відсутність потреби в системному адміністраторі.

Об'єктна модель, яку використовує ПЗ Елфа, описує структуру об'єктів, що становлять схему класифікації архівних документів, їх атрибути та взаємозв'язки з іншими об'єктами. Характерною особливістю об'єктної моделі ПЗ Елфа є відсутність заздалегідь визначених елементів класифікації. Для кожного елемента класифікації створюється окремий тип об'єкта. Кількість типів об'єктів у ПЗ Елфа необмежена,

точніше, обмежується лише розумною достатністю. Кожному типу об'єкта задаються атрибути, кількість яких також не обмежена. Між типами визначаються зв'язки. Прикладом такої об'єктної моделі може бути класична модель «фондоутворювач – фонд – комплекс – справа – тощо». Але в ПЗ Елфа зв'язки між типами можуть бути довільними. У загальному випадку об'єктна модель ПЗ Елфа може бути представлена у вигляді розрідженого орієнтованого безконтурного графа [10]. Об'єктна модель зберігається в БД ПЗ Елфа, а її структуру визначають фахівці архівної установи за допомогою графічного інтерфейсу ПЗ Елфа.

На рис. 1 ієрархія об'єктної моделі починається з архівної установи. Це лише приклад, реальну об'єктну модель автори створили на підставі науково-дослідної роботи [11] у співпраці з ЦДНТА України. Ніщо не заважає побудувати об'єктну модель таким чином, щоб ієрархія починалася з Державної архівної служби України. У цьому випадку у ПЗ Елфа буде обліковано архівні документи всіх архівних установ. Але побудова такої об'єктної моделі потребує окремої науково-дослідної роботи.



Рис. 1. Приклад можливої об'єктної моделі ПЗ Елфа

Повнотекстовий пошук [12] у ПЗ Елфа створено на основі повнотекстового індексу, який являє собою перелік згадок слів у атрибутах об'єктів і вказівок, в яких місцях вони зустрічаються. Слова для цього переліку відбираються за допомогою спеціальних словників, таких самих, які використовуються для перевірки орфографії. До БД ПЗ Елфа підключено український та російський словники. Залежно від даних повнотекстового індексу може бути обчислена релевантність – порівняльна міра відповідності тексту атрибута до запиту. Для простоти реалізації і швидкодії пошуку при індексуванні не всі атрибути беруть участь у повнотекстовому пошуку. Набір атрибутів визначають фахівці архівної установи.

У разі повнотекстового пошуку, коли атрибутом є розпізнаний текст цифрової копії архівного документа в електронному вигляді, а в БД наявна сама цифрова

копія, результати пошуку будуть показані на растровому зображенні цифрової копії з текстом, виділеним за допомогою електронного маркера.

Ще одна відмінність програми – це модульний принцип побудови, який зображено на рис. 2 у вигляді діаграми модулів.

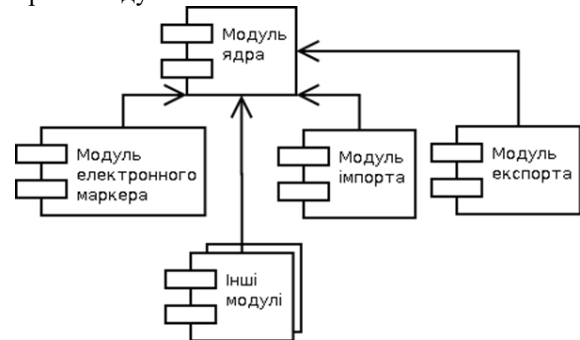


Рис. 2. Діаграма модулів ПЗ Елфа

Модульний принцип робить ПЗ Елфа універсальним для різних архівних установ. При використанні модульного принципу достатньо розробити додатковий програмний модуль для урахування особливостей предметної області кожної конкретної архівної установи. Модуль ядра при цьому залишається незмінним. У першу чергу це стосується модулів імпорту та експорту.

Модуль імпорту враховує той факт, що електронні облікові таблиці і бази даних у кожній архівній установі можуть мати свої форми та зберігатися в різних форматах, таких як MS Word, Excel тощо.

Модуль експорту враховує той факт, що архівні установи для публікації та експонування архівних документів можуть користуватися послугами різних Інтернет-провайдерів, які виставляють свої умови для розміщення документів у публічних електронних мережах.

Наразі ПЗ Елфа проходить дослідну експлуатацію на технічних засобах ЦДНТА України. Для забезпечення обліку та систематизації архівної інформації фахівцями ЦДНТА України здійснюється наповнення БД ПЗ Елфа. До БД вже внесено відомості стосовно:

- фондоутворювачів та історії їх перейменувань;
- фондів;
- комплексів.

Також до БД внесено зміст:

- описів (що наявні в електронному вигляді);
- частини систематичного каталогу;
- частини проектного каталогу;
- географічного каталогу;
- іменного каталогу.

Для спрощення наповнення БД у ПЗ Елфа передбачено операцію з імпорту архівних каталогів, описів тощо за умови, що вони зберігаються у форматах MS Word 97–2003 (.doc), MS Word 2007–2013 (.docx), MS Excel 97–2003 (.xls), MS Excel 2007–2013 (.xlsx), OpenDocument Text (.odt) та OpenDocument Spreadsheet (.ods). Після імпорту файли вищезазначених архівних каталогів та описів зберігаються у БД програми. Також до БД ПЗ Елфа завантажені наявні цифрові копії аркушів документів НАФ із розпізнаним

текстом. Це надає можливість здійснення пошуку не тільки за атрибутами облікових одиниць, а й за змістом архівних документів.

Повнофункціональна демонстраційна версія ПЗ Елфа доступна за посиланням: <http://micrography.gov.ua/info-res-adof/other/elfademo.zip>.

Висновки. Таким чином, упровадження ПЗ Елфа підвищить ефективність обліку архівних документів, придатних для експонування в електронному вигляді, та вторинної архівної інформації. Це у свою чергу ще на крок наблизить користувачів архівних установ до отримання вільного доступу до відомостей, потрібних для наукової діяльності. Підходи до побудови ПЗ Елфа дозволяють використовувати його в будь-якій архівній установі.

Список літератури

1. *Das Bundesarchiv*. URL: <https://www.bundesarchiv.de> (дата звернення: 16.03.2019).
2. *Світлова цифрова бібліотека*. URL: <https://www.wdl.org> (дата звернення: 16.03.2019).
3. *Документальні виставки on-line* / Державна архівна служба України. URL: <https://old.archives.gov.ua/Sections/Vystavky-\\region.php> (дата звернення: 16.03.2019).
4. *Виставки on-line* / Центральний державний науково-технічний архів України URL: <https://cdnta.archives.gov.ua/index.php/uk\\exhibitions-uk> (дата звернення: 16.03.2019).
5. *Великі дані*. URL: https://uk.wikipedia.org/wiki/Великі_дані (дата звернення: 16.03.2019).
6. *Перелік форматів даних електронних документів постійного і тривалого (понад 10 років) зберігання* : затверджено наказом Міністерства юстиції України від 11.11.2014 № 1886/5, зареєстровано в Міністерстві юстиції України від 11.11.2014 за № 1422/26199 / уклад. П. М. Марченко, Ю. С. Ковтанюк; Державна архівна служба України. Київ: УНДІАСД, 2011. 10 с.
7. Баранцев А. Ю. Поиск и отображение информации на растровых изображениях текстов, содержащихся в электронных документах. *Студії з архівної справи та документознавства* / Держ. архів. служба України. Київ: УНДІАСД, 2012. Т. 20, С. 136–139.
8. *Что такое PostgreSQL? Плюсы и минусы бесплатной базы данных*. URL: <https://oracle-patches.com/common/3214-что-такое-postgresql> (дата звернення: 16.03.2019).
9. Нестеров С. А. *Базы Данных*. Москва: Юрайт, 2017. 230 с.

10. Оре О. *Теория графов*. Москва: УРСС, 2008. 352 с.
11. *Дослідження перспектив застосування інтернет-технологій для забезпечення віддаленого доступу до документів НАФ, що зберігаються в ЦДНТА України, та довідкового апарату до них* / звіт про НДР (заключний); НДІ мікрографії; наук. кер. С. В. Ільїн. Харків, 2016. 68 с.
12. *Введение в полнотекстовый поиск в PostgreSQL*. URL: http://www.sai.msu.su/~megera/postgres/talks/fts_pgsql_intro.html (дата звернення: 16.03.2019).

References (transliterated)

1. *Das Bundesarchiv* URL: <https://www.bundesarchiv.de> (accessed 16.03.2019).
2. *Svitova cyfrova biblioteka*. URL: <https://www.wdl.org> (accessed 16.03.2019).
3. *Dokumental'ni vystavky on-line* / Derzhavna arhivna sluzhba Ukrainy. Kyi'v2017. URL: <https://old.archives.gov.ua/Sections\\Vystavky-region.php> (accessed 16.03.2019).
4. *Vystavky on-line* / Central'nyy derzhavnyy naukovy-technichnyy arhiv Ukrainy URL: <https://cdnta.archives.gov.ua/index.php/uk\\exhibitions-uk> (accessed 16.03.2019).
5. *Velyki_dani*. URL: https://uk.wikipedia.org/wiki/Великі_дані (accessed 16.03.2019).
6. *Perelik formativ danyh elektronnyh dokumentiv postynogo i tryvalogo (ponad 10 rokov) zberigannya*: zatverdzheno nakazom Ministerstva justycii' Ukrainy 11.11.2014 r., № 1886/5, zarejestrovano v Ministerstvi justycii' Ukrainy 11.11.2014 r. za № 1422/26199 / ukladachi: P. M. Marchenko, Ju. S. Kovtanjuk; Derzhavna arhivna sluzhba Ukrainy, UNDIASD. Kyi'v, 2011. 10 p.
7. A. Barancev. Poisk i otobrazhenie informacii na rastrovyyh izobrazheniyah tekstov, soderzhashhihsja v jelektronnyh dokumentah. *Studii' z arhivnoi' spravy ta dokumentoznavstva*. T. 20. Kyi'v, 2012. 136–139 p.
8. *Chto takoe PostgreSQL? Pljusy y minusy besplatnoj bazy dannyh*. URL: <https://oracle-patches.com/common/3214-что-такое-postgresql> (accessed 16.03.2019).
9. Nesterov S. A. *Bazy Danyh* [Database]. Moskva: Jurajt, 2017. 230 p.
10. O. Ore. *Teoryja grafov* [Graph theory]. Moskva: URSS, 2008. 352 p.
11. *Doslidzhennja perspektiv zastosuвання internet-tehnologij dlja zabezpechennja viddalenoogo dostupu do dokumentiv NAF, shho zberigajut'sja v CDNTA Ukrainy, ta dovidkovogo aparatu do nyh* / NDI mikrografii'; zvit pro NDR (zakljuchnyj); nauk. ker. S. V. Il'in. Harkiv, 2016. 68 p.
12. *Vvedenie v polnotekstovyy poisk v PostgreSQL*. URL: http://www.sai.msu.su/~megera/postgres/talks/fts_pgsql_intro.html (accessed 16.03.2019).

Надійшла (received) 21.05.2020

Відомості про авторів / Сведения об авторах / About the Authors

Баранцев Андрій Юрійович – Науково-дослідний, проектно-конструкторський та технологічний інститут мікрографії (НДІ мікрографії), науковий співробітник; м. Харків, Україна; e-mail: barantsev@micrography.gov.ua

Клименко Наталія Миколаївна – Науково-дослідний, проектно-конструкторський та технологічний інститут мікрографії (НДІ мікрографії), провідний інженер-програміст; м. Харків, Україна; e-mail: klimenko@micrography.gov.ua

Шевченко Ігор Андрійович – Науково-дослідний, проектно-конструкторський та технологічний інститут мікрографії (НДІ мікрографії), інженер-технолог I категорії; м. Харків, Україна; e-mail: shevchenko@micrography.gov.ua

Баранцев Андрей Юрьевич – Научно-исследовательский, проектно-конструкторский и технологический институт микрографии (НИИ микрографии), научный сотрудник; г. Харьков, Украина; e-mail: barantsev@micrography.gov.ua

Клименко Наталья Николаевна – Научно-исследовательский, проектно-конструкторский и технологический институт микрографии (НИИ микрографии), ведущий инженер-программист; г. Харьков, Украина; e-mail: klimenko@micrography.gov.ua

Шевченко Игорь Андреевич – Научно-исследовательский, проектно-конструкторский и технологический институт микрографии (НИИ микрографии), инженер-технолог I категории; г. Харьков, Украина; e-mail: shevchenko@micrography.gov.ua

Barantsev Andrii Yuriiovych – project-design and technological institute of micrography (SRI micrography), Research Officer; Kharkiv, Ukraine; e-mail: barantsev@micrography.gov.ua

Klymenko Nataliia Mykolaivna – project-design and technological institute of micrography (SRI micrography), leading software engineer; Kharkiv, Ukraine; e-mail: klimenko@micrography.gov.ua

Shevchenko Ihor Andriiovych – project-design and technological institute of micrography (SRI micrography), engineer-technologist of the I category; Kharkiv, Ukraine; e-mail: shevchenko@micrography.gov.ua